



TITLE:

# Genome-Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by High-Throughput Sequencing

AUTHOR(S):

Chandran, Anandhakumar; Syed, Junetha; Li, Yue; Sato, Shinsuke; Bando, Toshikazu; Sugiyama, Hiroshi

---

CITATION:

Chandran, Anandhakumar ...[et al]. Genome-Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by High-Throughput Sequencing. ChemBioChem 2016, 17(20): 1905-1910

ISSUE DATE:

2016-10-17

URL:

<http://hdl.handle.net/2433/230875>

RIGHT:

This is the accepted version of the following article: [Anandhakumar Chandran, Junetha Syed, Yue Li, Shinsuke Sato, Toshikazu Bando, Hiroshi Sugiyama. Genome - Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by High - Throughput Sequencing. ChemBioChem (2016), 17, 20, 1905-1910], which has been published in final form at <https://doi.org/10.1002/cbic.201600274>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.; The full-text file will be made open to the public on 19 October 2017 in accordance with publisher's 'Terms and Conditions for Self-Archiving'.; この論文は出版社版でありません。引用の際には出版社版をご確認ご利用ください。; This is not the published version. Please cite only the published version.

# Genome-Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by Utilizing the Power of High-Throughput Sequencing

Anandhakumar Chandran,<sup>[a]</sup> Junetha Syed,<sup>[a]</sup> Yue Li,<sup>[a]</sup> Shinsuke Sato,<sup>[b]</sup> Toshikazu Bando,<sup>[a]</sup> Hiroshi Sugiyama,<sup>\*,[a,b]</sup>

**Abstract:** One of the major goals in DNA-based personalized medicine is the development of sequence-specific small molecules to target the genome by means of synthetic biology; SAHA-PIPs belong to such class of small molecules. In a complex eukaryotic genome, the differential biological effects of SAHA-PIPs remain unclear. These questions can be addressed by directly identifying the binding regions of small molecules across the genome; however, it is a challenge to enrich specifically the small-molecule-bound DNA without chemical cross-linking. Here, we developed a method using high-throughput sequencing to map the binding area of non-cross-linked small molecules throughout the chromatinized human genome. Analysis of the sequenced data confirmed the presence of specific binding sites for SAHA-PIPs among the enriched sequence reads. Mapping the binding sites and enriched regions on the human genome clarifies the origin of the distinctive biological effects of SAHA-PIP. This approach will be useful for identifying the functionality of other small molecules on a large scale.

## Introduction

The competency of each cell type to maintain its precise biological characteristics depends on the inherited differences in chromatin packaging named as nucleosomes. These preset arrangements in the genome enable the epigenome to control the fate of the cell by regulating a specific set of genes.<sup>[1,2]</sup> Notably, changes in the epigenetic machinery can lead to cell plasticity, which triggers cellular reprogramming.<sup>[3]</sup> Although various histone marks are considered to be the gears of the epigenome, methylation and acetylation are the extensively studied histone modifications in governing chromatin dynamics.<sup>[4]</sup> Epigenetic alterations by these chromatin-modifying enzymes result in the alteration of gene expression. Given that most epigenetic marks are dynamic, some of these post translational histone modifications are reversible by enzymes.<sup>[5]</sup> Synthetic biology approaches aim at controlling these modifications by using sequence-specific artificial small molecules and is one of the key objectives in developing personalized nucleic acid targeted therapy.

The use of the versatile *N*-methylpyrrole (P) *N*-methylimidazole (I) synthetic polyamides (PIPs) in regulating

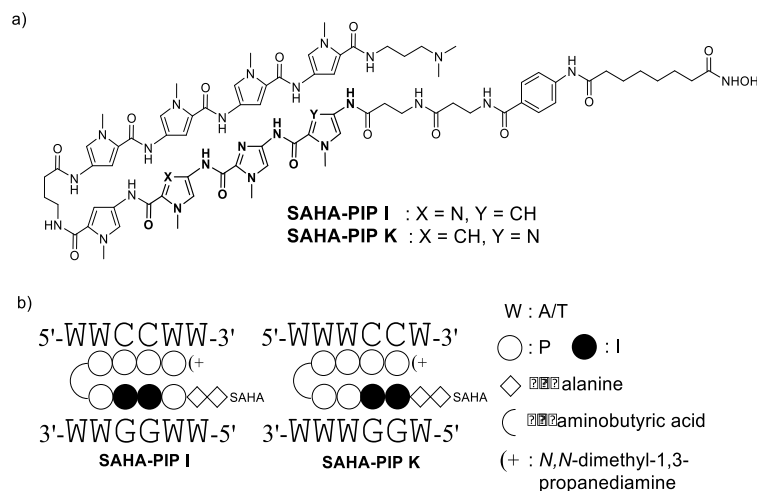
sequence specific gene expression has been successful and these molecules are efficient enough to penetrate cell membranes even at nanomolar concentrations. PIPs belong to the category of DNA minor-groove binders that can recognize each of the Watson–Crick base pairs with a programmable canonical DNA binding rule. The predetermined DNA binding rule of PIP is that an antiparallel arrangement of I opposite P (I–P) recognizes a G–C base pair, P–I recognizes a C–G base pair, and P–P recognizes either T–A or A–T base pairs.<sup>[6,7]</sup>

Previously, we have reported the design and application of SAHA-PIPs as epigenetically active artificial transcriptional regulators. SAHA-PIP was synthesized by conjugating a histone deacetylase inhibitor SAHA (suberoyl anilide hydroxamic acid) with PIPs. These studies showed that SAHA-PIPs can specifically trigger efficient transcriptional activation of an epigenetically silent gene network in human dermal fibroblasts.<sup>[8]</sup> However, the actual binding sites of SAHA-PIPs in a broad context were unclear. Therefore, previously we used Bind-n-Seq methods to understand the sequence recognition property of the SAHA-PIPs.<sup>[9]</sup> Our investigation on the global gene expression of a chemical library containing 32 SAHA-PIPs in human dermal fibroblast cells showed that each of the conjugates can regulate a unique gene set, including some therapeutically important genes.<sup>[10]</sup> The underlying factors and characteristics of SAHA-PIPs that lead to regulation of a unique gene set remain unresolved. Previous studies with alkylating PIP suggested that the genomic regions occupied by histones can influence the binding and gene regulating effect of PIPs.<sup>[11]</sup>

Beforehand, “Cross-linking of small molecules for isolation of chromatin” (COSMIC) was used to determine the high-affinity binding of PIP in the nucleus by conjugating a photo-cross-linker with PIP.<sup>[12]</sup> Our recent report on the use of alkylating-PIP was further extended to examine the binding conformation.<sup>[13]</sup> These studies established a platform upon which to investigate the PIP conjugate binding mechanism in chromatinized human genome. It remained an unresolved challenge to understand the binding behavior of SAHA-PIPs without any covalent cross-linking to DNA. In this report, we have developed a method to understand the binding behavior of SAHA-PIPs by combining micrococcal nuclease (MNase) digestion, affinity purification, and high-throughput sequencing. Our recent studies showed that SAHA-PIP I (Scheme 1) could precisely trigger the expression of essential pluripotency genes such as *OCT-3/4*, *NANOG*, and *DPPA4*, on the otherhand its structural counterpart, SAHA-PIP K (Scheme 1), activated a completely different set of genes related to germ-cell such as *PIWIL2*, *PIWIL4*, and *MOV10L1* in human dermal fibroblast cells.<sup>[14,15]</sup> These observations prompted us to study the genomic occupancy of SAHA-PIPs I and K on the chromatinized human genome. The results of this study reveal the high-affinity binding sites and binding preferences of SAHA-PIPs across the complex human genome.

[a] Dr. A. Chandran, Dr. J. Syed, Dr. Y. Li, Dr. T. Bando, Prof. H. Sugiyama  
Department of Chemistry, Graduate School of Science Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan.  
E-mail: hs@kuchem.kyoto-u.ac.jp

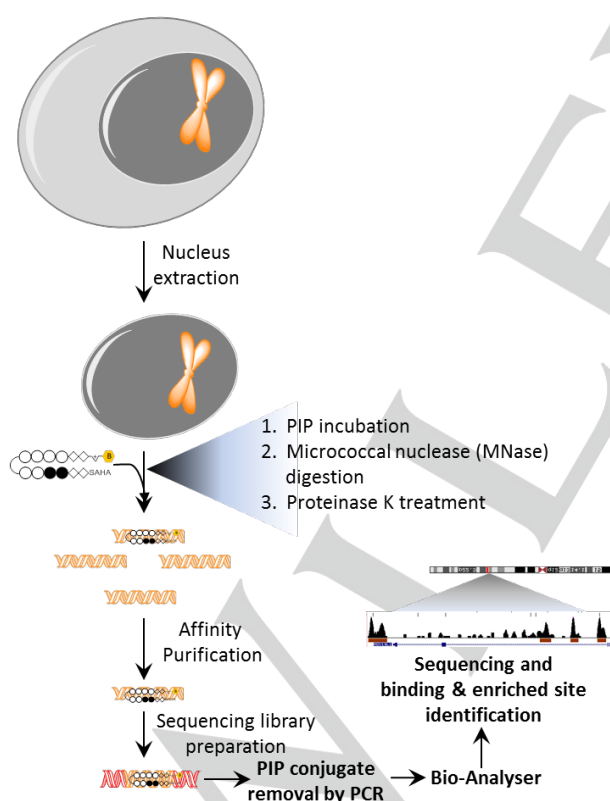
[b] S. Sato, Prof. H. Sugiyama  
Institute for Integrated Cell-Material Sciences (WPI-iCeMS) Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501 Japan.  
Supporting information for this article is given via a link at the end of the document.



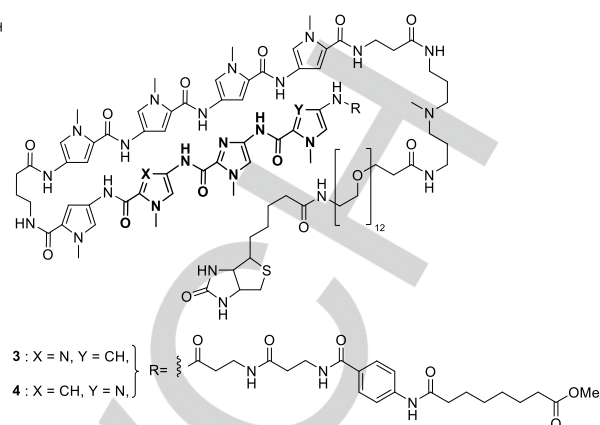
**Scheme 1.** a) Structure of SAHA-PIP I and SAHA-PIP K. b) SAHA-PIP I and SAHA-PIP K binding sequences based on PIP-DNA binding rule.

## Results and Discussion

As an artificial transcription activators, the biological effects of SAHA-PIP I and K mainly rely on their high-affinity binding preferences in the chromatinized genome. To characterize this phenomenon, we used our recent report<sup>[13]</sup> as a platform and developed a method shown in Figure 1, with non-covalently binding small molecules in the extracted nucleus of live cells.



**Figure 1.** MNase digestion and affinity purification-based high-throughput sequencing method pipeline using noncovalently binding PIP conjugates in the human nucleus from live cells.



**Scheme 2.** Chemical structures of **3** and **4**

For this study, we used SAHA-PIP I (SAHA- $\beta$ - $\beta$ -Py-Im-Im-Py- $\gamma$ -Py-Py-Py-Py- $\beta$ -(+)-PEG12-Biotin; **3**) and SAHA-PIP K (SAHA- $\beta$ - $\beta$ -Im-Im-Py-Py- $\gamma$ -Py-Py-Py-Py- $\beta$ -(+)-PEG12-Biotin; **4**), from our previously reported library of biotin-conjugated SAHA-PIPs<sup>[9]</sup> (Scheme 2). The compounds were synthesized as described in the previous report (see the Supporting Information). Our results unambiguously showed that neither the modification in the SAHA moiety nor attachment of biotin in SAHA-PIPs affected their binding specificity.<sup>[9]</sup>

In brief, the nucleus was extracted from live human fibroblast BJ cells (neonatal foreskin (ATCC, USA)) and incubated with **3** and **4** (400 nM) separately, control experiments were performed without SAHA-PIPs. To avoid the dissociation of non-covalently bound **3** and **4**, and to obtain the PIP bound target DNA fragment, MNase digestion was performed<sup>[16]</sup> (see the Supporting Information). After DNA fragmentation and nuclear protein digestion by proteinase K, PIP bound DNA was enriched and purified by using biotin-streptavidin chemistry based affinity purification. The extracted genomic DNA from the biological triplicates were used to construct sequencing libraries and subjected to high-throughput sequencing, using standard sequencing methods for Ion PGM<sup>TM</sup>/Proton<sup>TM</sup>.

High-quality sequencing reads were mapped with the human genome. To find the high-affinity binding sites of **3** and **4** over the control sequence reads (without any PIP treatment), uniquely mapped sequence data were randomly extracted by using subsampling Perl code and analyzed with a Bind-n-Seq pipeline and motif calling.<sup>[17]</sup> The identified high-affinity binding motifs for **3** and **4** are shown in Figure 2 and the sequence details are in Table S1-S6. One of the highly enriched motifs shown in figure 2a for **3** obeyed well its canonical binding rule (5'-WWCCWW-3'), but the palindromic nature of the sequence could not allow the easy determination of its binding orientation.

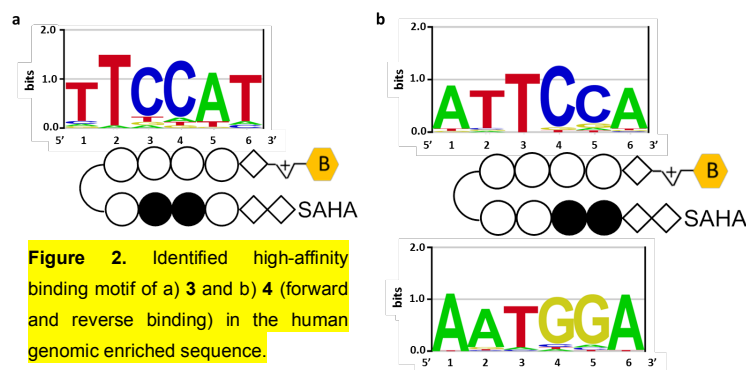


Figure 2b shows the high affinity binding sites of PIP conjugate **4** with significant enrichment scores; the obtained data revealed that **4** can bind to DNA in both the forward (N-terminal to C-terminal of PIP recognize 5' to 3' of DNA) and reverse orientation (N-terminal to C-terminal of PIP recognize 3' to 5' of DNA) with an enrichment ratio of 13.91 and 14.63, respectively. This type of binding was also observed in our previous report.<sup>[9]</sup> Comparative assessments of SAHA-PIPs binding on chromatinized genomic DNA with a protein-free synthetic DNA library<sup>[9]</sup> showed that both experimentally derived high-affinity binding motifs follow the PIP binding rule. Although the side-by-side arrangement of P-I-I-P in PIP recognizes a unique base pair (CG/GC), the P-P arrangement shows a slight variation in recognizing A/T (W). This variation may be due to the primary and secondary preferential nucleotide distribution in the given sequence context. Interestingly their base recognition(W) strictly follows the canonical binding rule.

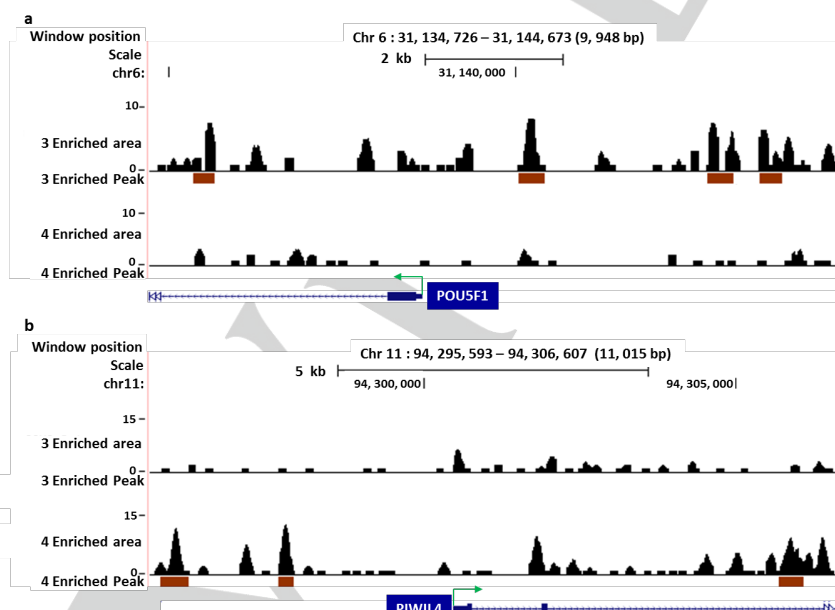
**Revealing the unique gene set activation mechanism of PIP conjugates in the human genome:** The epigenetically active form of **3** (SAHA-PIP I) (Scheme 1) was reported to activate an epigenetically silent pluripotency gene set in human fibroblast

cells,<sup>[14]</sup> whereas its structural counterpart epigenetically active form of **4** (SAHA-PIP K) (Scheme 1) triggered the activation of meiosis controlling PIWI pathway genes in somatic cells, which is involved in germ-cell generation.<sup>[15]</sup> When the obtained **3** and **4** enriched sequence read signal and enrichment peaks (peaks were identified with respect to control data)<sup>[18]</sup> were mapped, the results were noteworthy. The data for PIP conjugate **3** showed enrichment peaks at the promoter regions of SAHA-PIP I upregulated genes associated with pluripotency: *OCT3/4* (also known as *POU5F1*: POU domain, class 5, transcription factor 1) (Figure 3a), *DPPA4* (developmental pluripotency associated 4) (Figure S2a), and *EPCAM* (Figure S2b). In contrast, **4** did not show any significant level of enrichment at the SAHA-PIP I induced pluripotent genes. In line with this, enrichment peaks for **4** were observed in the promoter region and gene body of SAHA-PIP K upregulated genes involved in germ-cell generation such as *PIWIL4*, *PIWIL2*, and *TDRD9* (Figure 3b, S2c and S2d respectively). Careful observation of *TDRD9* shown in Figure S2d demonstrated the enrichment of peaks in the gene body but not around the promoter, explaining the reason behind the mild effect of SAHA-PIP K on *TDRD9* promoter acetylation, and the minor change in mRNA expression noted in our previous report.<sup>[15]</sup> Interestingly, compound **3** did not show any specific significant enrichment on either *PIWIL4* or *TDRD9*; even though it displayed a mild enhancement on *PIWIL2*, but it could not be comparable with the effective significant enrichment by conjugate **4**.

In contrast, neither **3** nor **4** showed any clearly enriched region on the housekeeping gene *GAPDH* (Figure S2e). Although the active forms of SAHA-PIP I and SAHA-PIP K have closely related recognition sites, their transcriptional activation network is distinctive to each other. Taken all together, our high-throughput sequencing analysis of **3** and **4** enriched genomic regions in chromatinized human genome allowed us to provide; the direct evidence of DNA minor-groove binding SAHA-PIP's differential gene activation mechanism in eukaryotic cells.

#### Identification of the preferential binding region of PIP conjugates in the tightly packed heterochromatin region:

As noted in the previous report, histone-occupied chromatin regions play an important role in PIP binding preferences in genomic DNA.<sup>[11]</sup> To study the binding inclinations of SAHA-PIPs, we compared the genomic regions that were enriched with **3** and **4** with the MNase-seq data of nucleosomal positioning from ENCODE at UCSC.<sup>[19,20]</sup> Heterochromatin play a critical role in gene silencing because of the tighter packaging of DNA.<sup>[21]</sup>





Covalent modifications of eukaryotic histone tails such as histone methylation are vital for heterochromatin formation;<sup>[21,22]</sup> these modifications occur mainly through trimethylation at histone 3 lysine 9 (H3K9)<sup>[21,23]</sup> and lysine 27 (H3K27).<sup>[23,24]</sup> We therefore analyzed and compared the H3K9me3 and H3K27me3

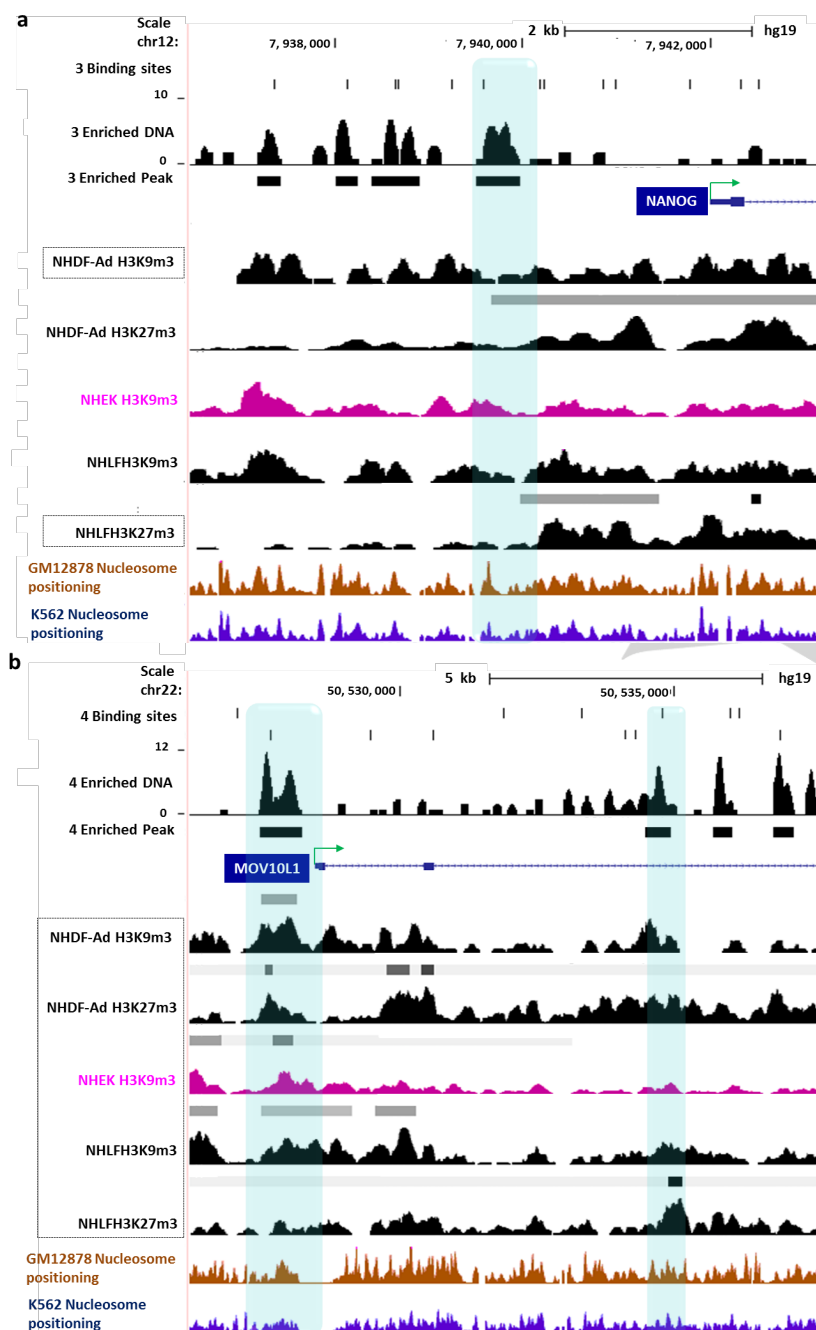
regions that were identified on the skin and lung fibroblast cell lines (NHEK (skin epidermal keratinocytes), NHDF-Ad (adult dermal fibroblasts skin), and NHLF (lung fibroblasts)),<sup>[20]</sup> along with the nucleosomal positioning data.

Our comparative analysis showed that the **3** enriched genomic region with high-affinity binding site around the SAHA-PIP I upregulated *NANOG*<sup>[14]</sup> promoter is H3K27 trimethylated (Shown in Figure 4a with blue shade) in the reported fibroblast cell lines and also falls under the nucleosomal region (MNase-seq data) (Figure 4a, last two tracks). Similarly, SAHA-PIP K highly upregulated *MOV10L1*<sup>[14]</sup> promoter region and gene body is enriched in **4**, (also harboring its binding site) are trimethylated at both H3K9 and H3K27 (Figure 4b, blue shade) in the reported fibroblast cell lines; these regions are also packed into nucleosomal units (MNase-seq data) (Figure 4b, last two tracks).

These results clearly demonstrate that both **3** and **4** can also bind efficiently to the nucleosome within the possible heterochromatin-forming regions. This provides evidence on silent gene activation principle of SAHA-PIPs.

## Conclusions

The application of high-throughput sequencing in chemical biology has produced some notable outcomes.<sup>[25]</sup> Furthermore, the identification of genomic targets by using small molecules help us in understanding the drug effects and optimise the drug design.<sup>[26]</sup> The combination of ChIP-seq and Chem-seq approaches<sup>[27,28]</sup> have been reported to characterise the protein-binding small molecules. Here, we report a method involving small molecules that form noncovalent interaction with DNA. We studied the high-affinity binding motif of such molecules and the corresponding enriched genomic regions. Annotation of **3** and **4** enriched region showed similar pattern of genomic region distribution (slightly varied in some genomic region) with high correlation among the experimentally identified peaks, actual binding sites and the deduced binding sites based on predetermined binding rule shown in Figure S3a and S3b. Comparative analysis of **3** and **4** provided evidence for the differential activation of the gene network by epigenetically active forms such as SAHA-PIP I and SAHA-PIP K. These results also suggest a binding mechanism for SAHA-PIPs in actual chromatinized genome. The following points can be noted in particular. i) Small variations in the P-I arrangement in PIP can result in a large difference in genome-wide binding site recognition. ii) The result that **3** and **4** bind on the nucleosomal region with histone marks may explain the epigenetic activation by SAHA-PIP of the silenced gene network. Given that this method



**Figure 4.** Binding and enrichment of PIP conjugates **3** and **4** in possible heterochromatin-forming regions of a) *NANOG* promoter and b) *MOV10L1* promoter. Possible heterochromatin regions were identified by mapping the MNase-Seq data of nucleosomal positioning with histone trimethylated regions of histone 3 lysine 9 and lysine 27. (Modified histone regions with PIP binding and enrichment site is shown in box.)

used non-cross-linking or noncovalent binding-based affinity purification, the approach is expected to be widely applicable to the study of genomic effects of other DNA-binding small molecules.

## Experimental Section

See the Supporting Information for Synthesis, experimental protocols and data analysis

## Acknowledgements

This work was partially supported by JSPS KAKENHI (grant number 24225005), Basic Science and Platform Technology Program for Innovative Biological Medicine from Japan Agency for Medical Research and Development, AMED, and JSPS-NSF International Collaborations in Chemistry (ICC) to H.S., and by JSPS KAKENHI (grant number 24310155) to T.B.

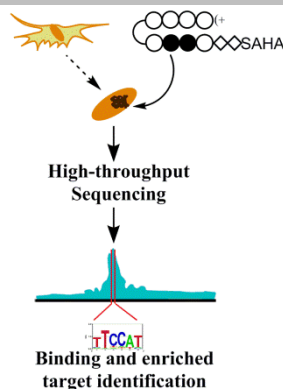
**Keywords:** SAHA-PIP • Pull-down • Epigenetics • Small molecule • High-throughput sequencing

## References:

- [1] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, *Cell* **2007**, *131*, 861–872.
- [2] F. Mohn, D. Schübeler, *Trends Genet.* **2009**, *25*, 129–136.
- [3] B. Pachaiyappan, P. M. Woster, *Bioorganic Med. Chem. Lett.* **2014**, *24*, 21–32.
- [4] C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, *Nat. Rev. Drug Discov.* **2012**, *11*, 384–400.
- [5] T. Kouzarides, *Cell* **2007**, *128*, 693–705.
- [6] C. D. Carlson, C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina, A. Z. Ansari, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4544–9.
- [7] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
- [8] a) A. Ohtsuki, M. T. Kimura, M. Minoshima, T. Suzuki, M. Ikeda, T. Bando, H. Nagase, K. ichi Shinohara, H. Sugiyama, *Tetrahedron Lett.* **2009**, *50*, 7288–7292; b) G. N. Pandian, K. I. Shinohara, A. Ohtsuki, Y. Nakano, M. Masafumi, T. Bando, H. Nagase, Y. Yamada, A. Watanabe, N. Terada, et al., *ChemBioChem* **2011**, *12*, 2822–2828; c) G. N. Pandian, A. Ohtsuki, T. Bando, S. Sato, K. Hashiya, H. Sugiyama, *Bioorganic Med. Chem.* **2012**, *20*, 2656–2660; d) G. N. Pandian, Y. Nakano, S. Sato, H. Morinaga, T. Bando, H. Nagase, H. Sugiyama, *Sci. Rep.* **2012**, *2*, 1–8.
- [9] C. Anandhakumar, Y. Li, S. Kizaki, G. N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, *ChemBioChem* **2014**, *15*, 2647–2651.
- [10] G. N. Pandian, J. Taniguchi, S. Junetha, S. Sato, L. Han, A. Saha, C. AnandhaKumar, T. Bando, H. Nagase, T. Vijayananthi, et al., *Sci. Rep.* **2014**, *4*, 3843.
- [11] C. Jespersen, E. Soragni, C. James Chou, P. S. Arora, P. B. Dervan, J. M. Gottesfeld, *Bioorganic Med. Chem. Lett.* **2012**, *22*, 4068–4071.
- [12] G. S. Erwin, D. Bhimsaria, A. Eguchi, A. Z. Ansari, *Angew. Chemie - Int. Ed.* **2014**, *53*, 10124–10128.
- [13] A. Chandran, J. Syed, R. D. Taylor, G. Kashiwazaki, S. Sato, K. Hashiya, T. Bando, H. Sugiyama, *Nucleic Acids Res.* **2016**, doi: 10.1093/nar/gkw283.
- [14] G. N. Pandian, S. Sato, C. Anandhakumar, J. Taniguchi, K. Takashima, J. Syed, L. Han, A. Saha, T. Bando, H. Nagase, et al., *ACS Chem. Biol.* **2014**, *9*, 2729–2736.
- [15] L. Han, G. N. Pandian, S. Junetha, S. Sato, C. Anandhakumar, J. Taniguchi, A. Saha, T. Bando, H. Nagase, H. Sugiyama, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 13410–3.
- [16] Published in association with Cold Spring Harbor Laboratory Press. *Nat. Meth.* **2005**, *2*, 719–720.
- [17] a) A. Zykovich, I. Korf, D. J. Segal, *Nucleic Acids Res.* **2009**, *37*, e151; b) T. L. Bailey, *Bioinformatics* **2011**, *27*, 1653–1659; c) C. T. Workman, Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, P. V. Benos, *Nucleic Acids Res.* **2005**, *33*, DOI 10.1093/nar/gki439.
- [18] J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, *Nat. Protoc.* **2012**, *7*, 1728–1740.
- [19] a) W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, a. D. Haussler, *Genome Res.* **2002**, *12*, 996–1006; b) D. Karolchik, G. P. Barber, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haussler, et al., *Nucleic Acids Res.* **2014**, *42*, DOI 10.1093/nar/gkt1168; c) K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haussler, et al., *Nucleic Acids Res.* **2015**, *43*, D670–D681.
- [20] a) Encode Consortium, *Nature* **2013**, *489*, 57–74; b) J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shores, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, et al., *Nature* **2011**, *473*, 43–9.
- [21] S. I. S. Grewal, J. C. Rice, *Curr. Opin. Cell Biol.* **2004**, *16*, 230–238.
- [22] C. Martin, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–849.
- [23] A. H. F. M. Peters, S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. H. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, et al., *Mol. Cell* **2003**, *12*, 1577–1589.
- [24] A. Santenard, C. Ziegler-Birling, M. Koch, L. Tora, A. J. Bannister, M.-E. Torres-Padilla, *Nat. Cell Biol.* **2010**, *12*, 853–862.
- [25] C. Anandhakumar, S. Kizaki, T. Bando, G. N. Pandian, H. Sugiyama, *ChemBioChem* **2015**, *16*, 20–38.
- [26] R. Rodriguez, K. M. Miller, *Nat Rev Genet* **2014**, *15*, 783–796.
- [27] L. Anders, M. G. Guenther, J. Qi, Z. P. Fan, J. J. Marineau, P. B. Rahl, J. Lovén, A. A. Sigova, W. B. Smith, T. I. Lee, et al., *Nat. Biotechnol.* **2014**, *32*, 92–96.
- [28] C. Jin, L. Yang, M. Xie, C. Lin, D. Merkurjev, J. C. Yang, B. Tanasa, S. Oh, J. Zhang, K. A. Ohgi, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9235–40.

## FULL PAPER

**Evaluation of artificial transcriptional activating small molecules binding effects on human genome.** Our previously reported SAHA PIPs library showed unique gene set activation. But its gene activation mechanism is uncertain with a complex human genome. Therefore, we developed high-throughput sequencing based method to address this query with a noncross-linking small molecule.



A. Chandran, J. Syed, Y. Li, S. Sato, T. Bando, H. Sugiyama\*

Page No. – Page No.

**Genome-Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by Utilizing the Power of High-Throughput Sequencing.**

## Supporting information

### **Genome-Wide Assessment of the Binding Effects of Artificial Transcriptional Activators by Utilizing the Power of High-Throughput Sequencing**

*Anandhakumar Chandran, Junetha Syed, Yue Li, Shinsuke Sato, Toshikazu Bando, Hiroshi  
Sugiyama,\**



## Table of contents

Experimental section.....	2
Supporting Figures.....	7
Supporting Table.....	11



Solvents and reagents were used as same as our previous report.<sup>[1]</sup> The EZ-Link NHS-PEG<sub>12</sub>-Biotin was purchased from Thermo Scientific (number 21312). The analytical HPLC was performed with a COSMOSIL 5C18-MS-II reversed phase column (4.6×150 mm, Nacalai) in 0.1% TFA in water with CH<sub>3</sub>CN as eluent at a flow rate of 1.0 mL/min, and a linear gradient elution of 0–100% CH<sub>3</sub>CN over 40 min with detection at 254 nm. The final products were analyzed by ESI-TOF-MS (Bruker).

We followed the same synthetic route as our previous report.<sup>[1]</sup> PIP supported by oxime resin was prepared in a stepwise reaction by a reported Fmoc solid-phase procedure. The product with oxime resin was cleaved with 3,3'-Diamino-*N*-methyldipropylamine (500 μL) at 55°C for 3 hours. After filtration and evaporation, the resulted oil was quenched by Et<sub>2</sub>O. The obtained precipitation was washed with Et<sub>2</sub>O for three times and dried in vacuo. The product was used for the next coupling steps without further purification.

A solution of polyamide (1eq), EZ-Link NHS-PEG<sub>12</sub>-Biotin (1.2eq) and DIEA (2eq) in DMF was stirred at room temperature for 1 hour. After consumption of starting material was confirmed by HPLC, Et<sub>2</sub>O was added to the mixture and the resultant was collected by centrifugation, and washed by Et<sub>2</sub>O and CH<sub>2</sub>Cl<sub>2</sub>. The crude product was purified by reverse-phase HPLC. After lyophilization the final product was obtained.

### **SAHA-β-β-Py-Im-Im-Py-γ-Py-Py-Py-Py-β-(+)-PEG<sub>12</sub>-Biotin (3)**

Analytical HPLC: *t*<sub>R</sub> = 17.2 min (0.1% TFA-CH<sub>3</sub>CN, 0–100%, 40min). ESI-TOF-MS *m/z*: calcd for C<sub>119</sub>H<sub>173</sub>N<sub>29</sub>O<sub>31</sub>S [M+2H]<sup>2+</sup> 2538.2573, found 2538.0851.

### **SAHA-β-β-Im-Im-Py-Py-γ-Py-Py-Py-Py-β-(+)-PEG<sub>12</sub>-Biotin (4)**

Analytical HPLC: *t*<sub>R</sub> = 17.3 min (0.1% TFA-CH<sub>3</sub>CN, 0–100%, 40min). ESI-TOF-MS *m/z*: calcd for C<sub>119</sub>H<sub>173</sub>N<sub>29</sub>O<sub>31</sub>S [M+2H]<sup>2+</sup> 2538.2573, found 2538.2536.

## **Compound 3 and 4 binding site and binding region identification in human genome (Mainly we have followed our recent report)<sup>[2]</sup>**

### **Nucleus extraction and PIP conjugate incubation:**

75% -80% confluent human fibroblast BJ cells (neonatal foreskin (ATCC, USA)) were used to extract nuclei.<sup>[3]</sup> In brief, P6 of BJ-HFF cells were grown in Dulbecco's modified eagle medium (DMEM, Nacalai Tesque, Japan) supplemented with 10% FBS (FBS, Japan Serum) in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C. At 75%-80% confluency, the cells were isolated by PBS washing followed by 3-4 minutes trypsinization. Isolated 2x10<sup>6</sup> cells were washed twice with ice cold PBS. The cell membrane was digested with 5ml of ice cold NP-40 lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.5% Nonidet P-40, 0.15 mM spermine, 0.5 mM spermidine and 0.1x protease inhibitor cocktail) for 5 minutes on ice. Loosen pellet of cell nuclei were obtained by centrifugation and dissolved in binding and resuspension buffer (10 mM Tris-Cl (pH 8.0), 5 mM MgCl<sub>2</sub>, 1 mM DTT, 0.3 M KCl, 0.3x protease inhibitor cocktail and 10% glycerol). **3** and **4** were dissolved in DMSO and incubated (400nM of **3** and **4**, 0.1% final concentration of DMSO) separately with the isolated nuclei about 16-18 hrs at 4°C. For the control experiments only DMSO (PIPs were dissolved in DMSO) was used without PIP conjugates. Concentrations were used according to the previous reports.<sup>[2,4,5]</sup>

### **MNase digestion:**

After PIP derivatives incubation, micrococcal nuclease (MNase) buffer<sup>[3]</sup> (10 mM Tris-HCl (pH 7.4), 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine and 0.1x protease inhibitor cocktail) pre-washed nuclei was subjected to MNase (Takara, Japan) digestion at 37°C for 30 min. Sonication shearing may affect the non-covalently

bound PIP, so we used MNase digestion. To avoid protein hindrance during affinity purification, the MNase digested nucleosomes were treated with proteinase K.

### **Affinity purification**

Protein digested PIP bound DNA containing suspension was mixed with mixed with modified COSMIC buffer<sup>[4]</sup> (20 mM Tris-Cl (pH 8.1), 2 mM EDTA, 150 mM NaCl, 0.1x protease inhibitor cocktail, 1% Triton-X100, and 0.1% SDS). 10% of the sample was saved as input DNA. Dynabeads MyOne C1 preparation: Streptavidin-coated magnetic beads (Dynabeads MyOne C1, Life Technologies, USA) was washed twice with modified COSMIC buffer after removing the suspension solution. 0.5mg per sample of beads was incubated with samples at 4°C for 16 hrs by way of rotor mixing. After incubation, samples were subjected to sequential of washing with buffer 1, washing Buffer 2 (10 mM Tris-Cl (pH 8.0), 250 mM LiCl, 1 mM EDTA, 0.5% NP40), washing Buffer 3 (10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 0.1% NP-40), and finally with TE. Washed samples were resuspended in elution buffer-I (10 mM Tris-HCl (pH 7.6), 0.4 mM EDTA and 100 mM KOH) and DNA was eluted by heating at 90°C for 30 min. The unrecovered DNA with the beads were subjected to second elution using elution buffer-II (2% SDS, 100mM NaHCO<sub>3</sub> and 3mM biotin) with the heating of 65°C for 8-12 hours. The eluted DNA samples were purified with the QIAquick PCR purification Kit (Qiagen, CA, USA) and quantified using Nanodrop (Figure S1). The quantification of purified DNA samples confirmed the pull-down efficiency of **3** and **4** with respect to the biological triplicates (Figure S1).

### **High-throughput sequencing library construction and sequencing**

Optimum amount of purified DNA was pooled (To get sufficient amount of DNA) from



the biological triplicates and sequencing libraries were prepared by using Ion Xpress<sup>TM</sup> Plus gDNA Fragment Library preparation reagents and protocols (Life technologies, USA) as per the instruction. Adapter ligated DNA was amplified and purified. The purified libraries were analyzed with Agilent DNA High sensitivity BioAnalyzer kit (Agilent technologies, USA). The sequencing was carried out, as 1) template preparation using Ion PGM<sup>TM</sup> template OT2 200 v2 kit and Ion PI<sup>TM</sup> template OT2 200 kit in Ion one touch2 system. 2) The templates enrichment on Ion one touch ES. 3) Sequencing the enriched libraries with Ion PGM<sup>TM</sup>/ Ion Proton<sup>TM</sup> sequencer using Ion PGM<sup>TM</sup> sequencing 200 kit v2 / Ion PI<sup>TM</sup> Sequencing 200 kit v3 and 318 v2 chip/ Ion PI chip according to the manufacturer's guidelines. Single read sequencing was performed with 260-300 flow, 28-31 million post filtered reads per library was produced. Ion torrent suit was used for the preliminary data analysis. TMAP 4.4.2 was used for aligning the good quality reads with human genome, To identify high-affinity binding motif a random selected 10-15% of uniquely mapped reads were used (Normalization performed with control data). We followed our previous analysis pipeline for motif calling.<sup>[1, 2, 6]</sup> The aligned data was further analysed for enriched peak calling using standard ChIP-seq analysis program MACS 1.4.2.<sup>[7]</sup> (SAHA PIP enriched reads as a treat and control (without SAHA-PIP) reads as control) MACS enriched signals and peak regions were visualized on UCSC genome browser to identify uniquely enriched regions between 3 and 4.<sup>[8]</sup> To analyse the influence of chromatin bound DNA on PIP conjugates binding, we compared various histone marks and nucleosomal positioning ENCODE data with our obtained data.<sup>[9]</sup>

Figure S1

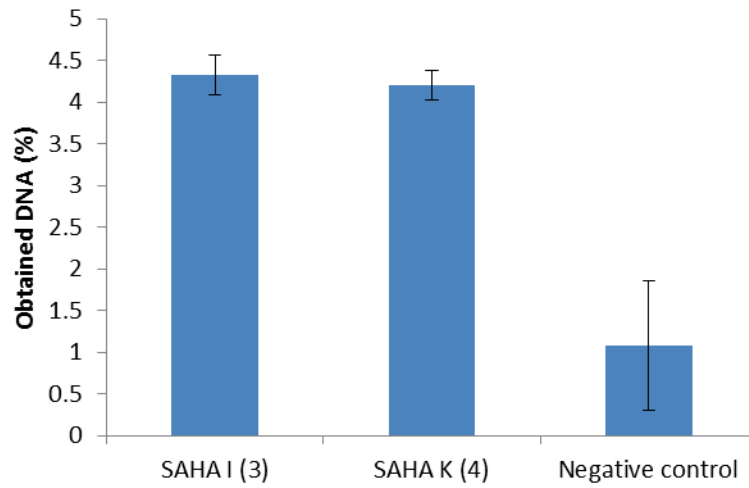
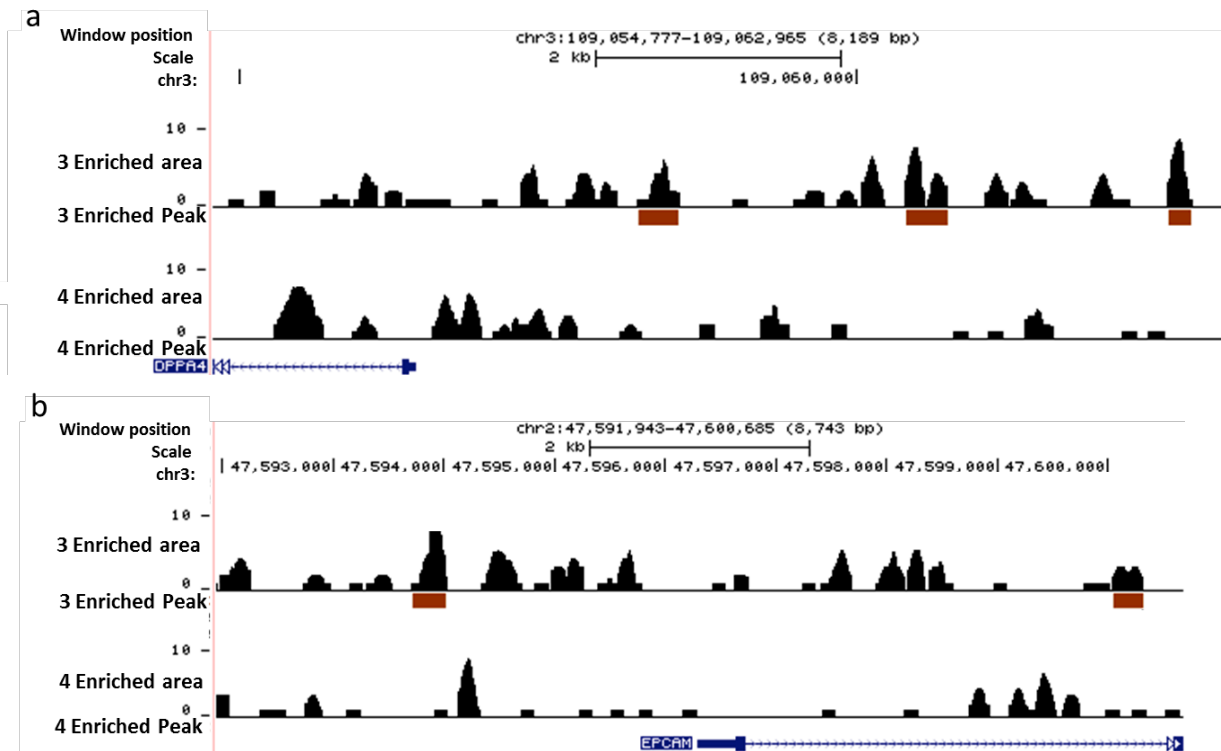
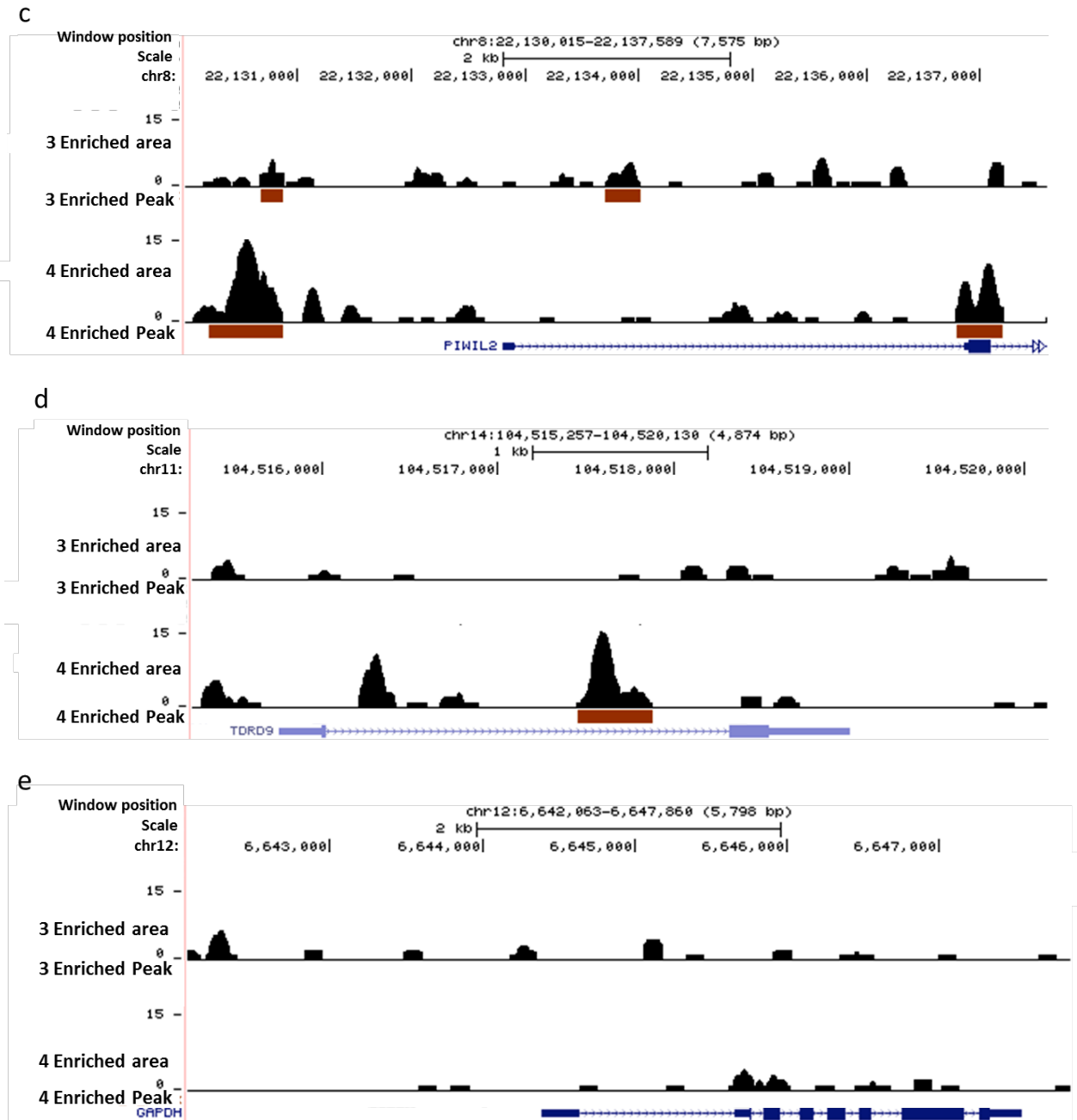


Figure S1. The quantification of purified DNA samples from the pull-down of 3 and 4.

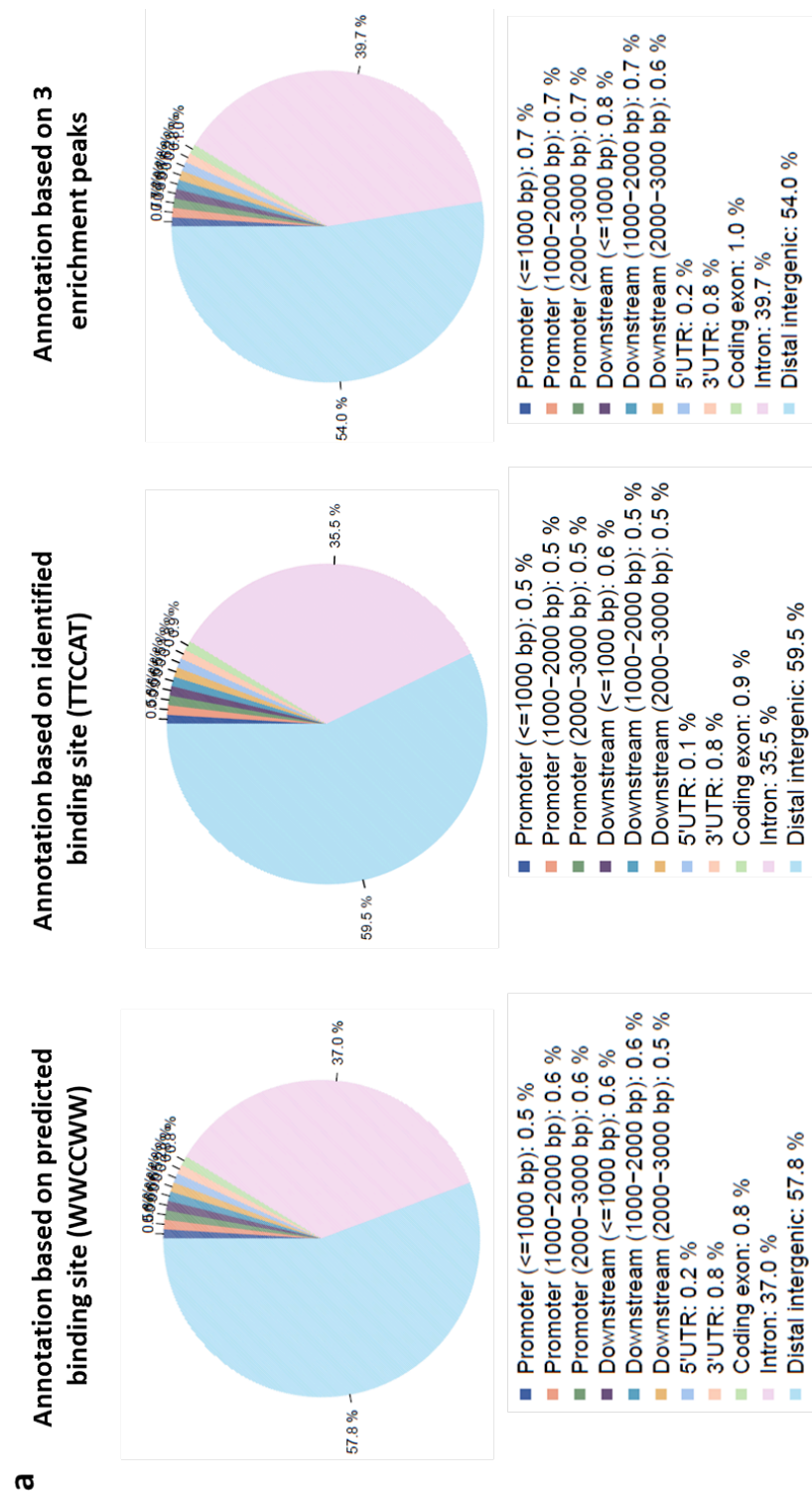
Figure S2

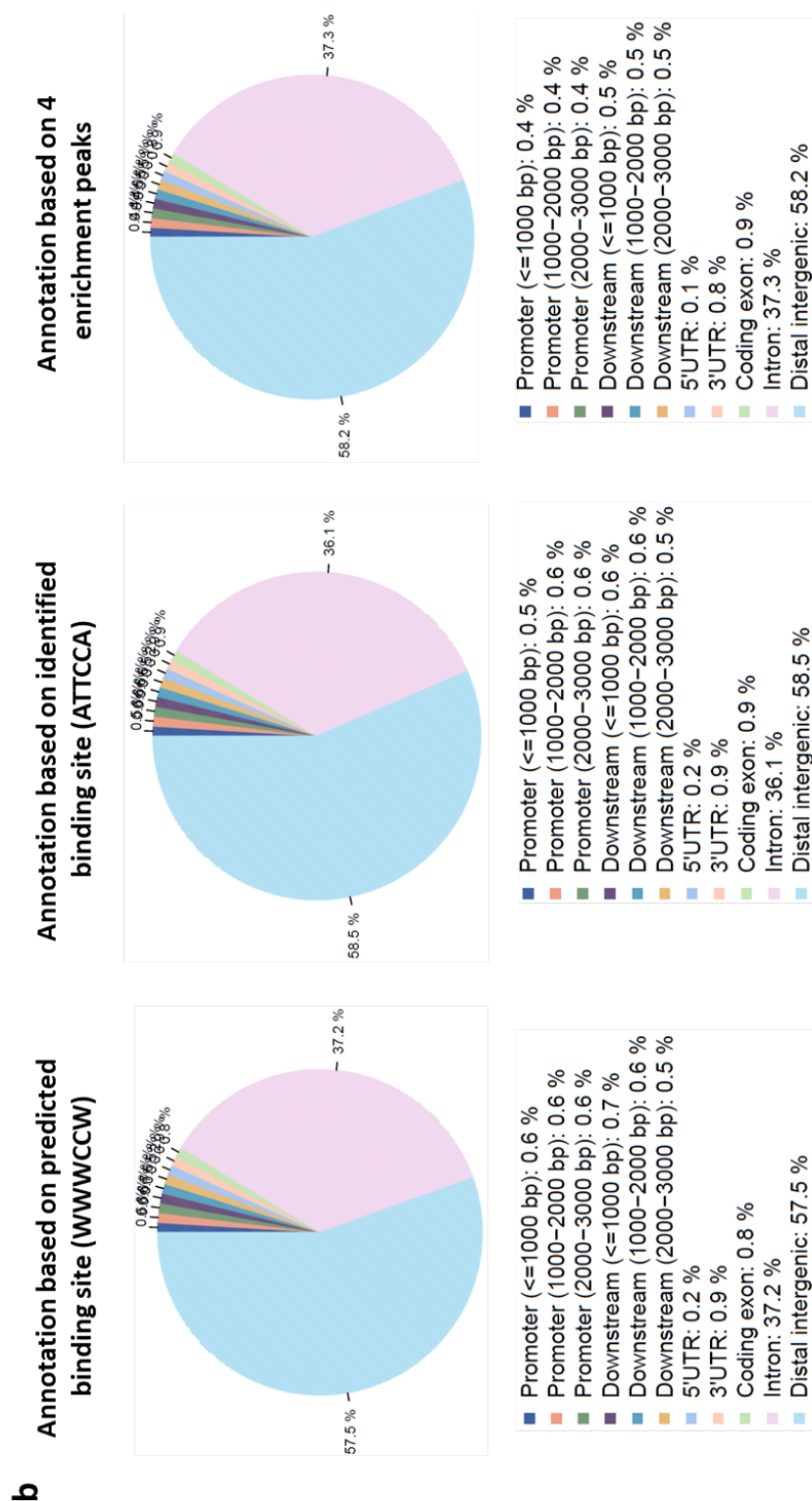




**Figure S2.** Identified genomic regions of PIP conjugates **3** and **4** binding and enrichment in a) *DPPA4* b) *EPCAM* c) *PIWIL2* d) *TDRD9* e) *GAPDH*.

Figure S3





**Figure S3.** Comparison of genomic binding and enriched region with binding rule based binding site. a) For compound **3**. b) For compound **4**.



**Table S1:** Base composition for **3** enriched high-affinity binding motif (Figure 2a)

Base/Position	1	2	3	4	5	6
A	101542	74463	95607	121875	1795588	135205
C	133568	0	1660957	1729644	0	128146
G	82898	0	110992	80128	44857	0
T	1723257	1966802	173709	109618	200820	1777914

**Table S2:** Base composition in percentage (%) for **3** enriched high-affinity binding motif (Figure 2a)

Base/Position	1	2	3	4	5	6
A	4.97	3.64	4.68	5.97	87.96	6.62
C	6.54	0	81.37	84.73	0	6.27
G	4.06	0	5.43	3.92	2.19	0
T	84.42	96.35	8.50	5.37	9.84	87.10

**Table S3:** Base composition for **4** enriched high-affinity forward binding motif (Figure 2b)

Base/Position	1	2	3	4	5	6
A	11484904	377658	154445	0	596910	11407282
C	0	449598	0	11554127	10168801	222343
G	330252	400567	0	476424	1186363	325987
T	619941	11207274	12280652	404546	483023	479485

**Table S4:** Base composition in percentage (%) for **4** enriched high-affinity forward binding motif (Figure 2b)

Base/Position	1	2	3	4	5	6
A	92.36	3.037	1.24	0	4.80	91.73
C	0	3.61	0	92.92	81.78	1.78
G	2.65	3.22	0	3.83	9.54	2.62
T	4.98	90.13	98.76	3.25	3.88	3.855

**Table S5:** Base composition for **4** enriched high-affinity reverse binding motif (Figure 2b)

Base/Position	1	2	3	4	5	6
A	10409874	9781071	646082	392937	521859	10299390
C	0	0	340220	659086	435721	0
G	0	481514	0	9062792	9343911	0
T	148100	295389	9571672	443159	256483	258584

**Table S6:** Base composition in percentage (%) for **4** enriched high-affinity reverse binding motif (Figure 2b)

Base/Position	1	2	3	4	5	6
A	98.60	92.64	6.11	3.72	4.94	97.55
C	0	0	3.22	6.24	4.13	0
G	0	4.56	0	85.84	88.50	0
T	1.40	2.80	90.66	4.20	2.430	2.45

## References:

1. C. Anandhakumar, Y. Li, S. Kizaki, G.N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, *Chembiochem* **2014**, *15*, 2647-2651.
2. A. Chandran, J. Syed, R. D. Taylor, G. Kashiwazaki, S. Sato, K. Hashiya, T. Bando, H. Sugiyama, *Nucleic Acids Res.* **2016**, doi: 10.1093/nar/ gkw283.
3. Published in association with Cold Spring Harbor Laboratory Press, *Nat. Meth.* **2005**, *2*, 719–720.
4. G. S. Erwin, D. Bhimsaria, A. Eguchi, A. Z. Ansari, *Angew. Chemie - Int. Ed.* **2014**, *53*, 10124–10128.
5. C. F. Hsu, P. B. Dervan, *Bioorganic Med. Chem. Lett.* **2008**, *18*, 5851–5855.
6. a) A. Zykovich, I. Korf, D. J. Segal, *Nucleic Acids Res.* **2009**, *37*, e151; b) T. L. Bailey, *Bioinformatics* **2011**, *27*, 1653–1659; c) C. T. Workman, Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, P. V. Benos, *Nucleic Acids Res.* **2005**, *33*, DOI 10.1093/nar/gki439.
7. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, *Nat. Protoc.* **2012**, *7*, 1728-1740.

8. a) W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, a. D. Haussler, *Genome Res.* **2002**, *12*, 996–1006; b) D. Karolchik, G. P. Barber, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, et al., *Nucleic Acids Res.* **2014**, *42*, DOI 10.1093/nar/gkt1168; c) K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, et al., *Nucleic Acids Res.* **2015**, *43*, D670–D681.
9. a) Encode Consortium, *Nature* **2013**, *489*, 57–74; b) J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shores, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, et al., *Nature* **2011**, *473*, 43–9.